Claims

1. A method for estimating statistics on an attribute of a relation, comprising

forming a histogram of said attribute of said relation, the histogram being augmented to identify the most frequent values of an attribute,

evaluating said histogram in connection with a criterion for retrieval of data from a relation.

2. The method of claim 1 further comprising forming a second histogram of said attribute of a second relation, said second histogram being augmented to identify the most frequent values of said attribute, and

evaluating said histograms to identify frequent values shared by said histograms..

3. The method of claim 2, further comprising multiplying frequent values in each of said histograms to produce a estimate of join fanout of a join of said relations on said attribute.

4. The method of claim 3, further comprising multiplying a number of a frequent value in one said histogram by an estimate of the average number infrequent values in the other histogram.

5. The method of claim 3 further comprising computing a number of matching infrequent values in each said histogram by

estimating a number of infrequent values in each relation using said histograms, and

computing from said estimates the join fanout attributable to said attribute.

6. A computer system for implementing a relational database system and performing a user query on said relational database system, comprising

storage for relations of said relational database system, and a histogram of an attribute of a first of said relations, the histogram being augmented to identify the most frequent values of an attribute in said first relation,

a computing circuit for implementing said relational database system, said computing circuit computing a statistic on said attribute by evaluating said histogram in connection with a criterion for retrieval of data from a relation.

7. The computer system of claim 6 wherein

said storage further includes a second histogram of said attribute of a second of said relations, said histogram being augmented to identify the most frequent values of said attribute in said second relation, and

said computing circuit evaluates said histograms to identify frequent values shared by said histograms..

8. The computer system of claim 7 wherein

said computing circuit multiplies frequent values in each of said histograms to produce a estimate of join fanout of a join of said relations on said attribute.

9. The computer system of claim 8 wherein

said computing circuit multiplies a number of a frequent value in one said histogram by an estimate of the average number infrequent values in the other histogram.

10. The computer system of claim 8 wherein

said computer system further computes a number of matching infrequent values in each said histogram by estimating a number of infrequent values in each relation using said histograms, and computing from said estimates the join fanout attributable to said attribute.

11. A program product for estimating statistics on an attribute of a relation, comprising

a program of instructions executable on a computer system to form a histogram of said attribute of said relation, the histogram being

augmented to identify the most frequent values of an attribute, and evaluate said histogram in connection with a criterion for retrieval of data from a relation, and

a signal bearing medium bearing the program.

12. The program product of claim 11 wherein said program further comprises instructions for forming a second histogram of said attribute of a second relation, said second histogram being augmented to identify the most frequent values of said attribute, and evaluating said histograms to identify frequent values shared by said histograms.

13. The program product of claim 12, wherein said program further comprises instructions for multiplying frequent values in each of said histograms to produce a estimate of join fanout of a join of said relations on said attribute.

14. The program product of claim 13, wherein said program further comprises instructions for multiplying a number of a frequent value in one said histogram by an estimate of the average number infrequent values in the other histogram.

15. The program product of claim 13 wherein said program further comprises instructions for computing a number of matching infrequent values in each said histogram by

estimating a number of infrequent values in each relation using said histograms, and

computing from said estimates the join fanout attributable to said attribute.

16. The program product of claim 11 wherein said signal bearing medium is a recordable medium.

17. The program product of claim 11 wherein said signal bearing medium is a transmission-type medium.